JOURNAL OF MULTIVARIATE ANALYSIS 41, 117–131 (1992)

Inadmissibility of Non-Order-Preserving Orthogonally Invariant Estimators of the Covariance Matrix in the Case of Stein's Loss

YO SHEENA AND AKIMICHI TAKEMURA

University of Tokyo, Tokyo, Japan

Communicated by the Editors

For orthogonally invariant estimation of Σ of Wishart distribution using Stein's loss, any estimator which does not preserve the order of the sample eigenvalues is dominated by a modified estimator preserving the order.

1. Introduction

Let $\mathbf{W}$ be distributed to $\mathbf{W}_p(k, \Sigma)$. Let

$$\mathbf{W} = \mathbf{HDH}'$$

be the spectral decomposition of $\mathbf{W}$, where $\mathbf{D} = \mathrm{diag}(l_1, ..., l_p)$, l_i is an eigenvalue of $\mathbf{W}$ $(0 \leqslant l_p \leqslant \cdots \leqslant l_1)$, and $\mathbf{H} = (h_{ij})$ is an orthogonal matrix. We consider estimation of Σ. We restrict our attention to orthogonally invariant estimators. An orthogonally invariant estimator can be written as

$$\hat{\mathbf{\Sigma}} = \mathbf{H\Psi H}', \qquad \mathbf{\Psi} = \mathrm{diag}(\psi_1(\boldsymbol{l}), ..., \psi_p(\boldsymbol{l})), \qquad \psi_i(\boldsymbol{l}) > 0 \ (1 \leqslant \forall i \leqslant p), \tag{1.1}$$

where $\boldsymbol{l} = (l_1, ..., l_p)$.

We use the Stein's loss, i.e.,

$$L(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) = \mathrm{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) - \log|\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}| - p. \tag{1.2}$$

Since the risk of an orthogonally invariant estimator depends only on the eigenvalues of $\mathbf{\Sigma}$, we can assume without loss of generality that

$$\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, ..., \sigma_p^2) \qquad (\sigma_1^2 \geqslant \cdots \geqslant \sigma_p^2).$$

Received November 7, 1990; revised May 15, 1991.

AMS 1980 subject classifications: 62C15, 62H12.

Key words and phrases: Wishart distribution, orthogonally invariant estimator, Stein's loss, inadmissibility, order-preserving.

0047-259X/92 $3.00

We also consider a subfamily of orthogonally invariant estimators. Suppose ψ_i is given by

$$\psi_i(\boldsymbol{l}) = d_i l_i, \tag{1.3}$$

where d_i is a positive constant. This family includes the unbiased estimator $d_1 = \cdots = d_p = 1/k$.

The unbiased estimator is not minimax and is dominated by the best invariant estimator $\hat{\boldsymbol{\Sigma}}_T$ with respect to the triangular group (see James and Stein [1]). $\hat{\boldsymbol{\Sigma}}_T$ is known to be minimax. Later Stein [2] and Dey and Srinivasan [5] showed that the orthogonally invariant estimator,

$$\psi_i(\boldsymbol{l}) = \delta_i l_i, \qquad \delta_i = \frac{1}{k+1+p-2i}, \tag{1.4}$$

dominates $\hat{\boldsymbol{\Sigma}}_T$. Dey and Srinivasan [5] also showed an estimator which further dominates (1.4) when $p \geqslant 3$. Their ψ_i is given by

$$\psi_i(\boldsymbol{l}) = \delta_i l_i - (l_i \log l_i)\, \tau(u)/(b_1 + u), \tag{1.5}$$

where $u = \sum_{i=1}^{p} \log^2 l_i$, $b_1 \geqslant 144(p-2)^2/25(k+p-1)^2$, $\tau(u)$ is a non-decreasing function which satisfies $0 < \tau(u) < 12(p-2)/5(k+p-1)^2$ and $E[\tau'(u)] < \infty$.

Here the following questions remain:

1. We call (1.1) order-preserving if the condition

$$\psi_1 \geqslant \psi_2 \geqslant \cdots \geqslant \psi_p \qquad \forall \boldsymbol{l}$$

is satisfied. Equations (1.4) and (1.5) are not order-preserving. This seems somewhat unnatural. Are there any general methods to improve non-order-preserving estimators? Isotonic regression (see Lin and Perlman [6]) is one method, but the improvement by the isotonic regression has not been proved theoretically. We will propose two methods of improving non-order-preserving estimators. One is the method using order statistics and the other is the isotonic regression. The improvement by these methods will be theoretically proved in Propositions 1 and 2.

2. When $p = 2$, it has not been proved that (1.4) is inadmissible. More generally, are the estimators of the form (1.3) inadmissible? Corollary 1 below answers this question.

2. Order-Preserving Estimator

Lemma 1. *Let*

$$a_j = (\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H})_{jj} = \sum_{m=1}^{p} h_{mj}^2 \sigma_m^{-2}.$$

Then for $1 \leqslant \forall s < \forall i \leqslant p$,

$$\begin{aligned} E_\Sigma[a_s \mid \boldsymbol{l}] &= E_\Sigma[a_i \mid \boldsymbol{l}] \qquad \forall \boldsymbol{l} \in \mathscr{L} \quad \textit{if } \mathbf{\Sigma} = c\mathbf{I} \\ E_\Sigma[a_s \mid \boldsymbol{l}] &< E_\Sigma[a_i \mid \boldsymbol{l}] \qquad \forall \boldsymbol{l} \in \mathscr{L} \quad \textit{otherwise}, \end{aligned} \tag{2.1}$$

where $E_\Sigma[a_s \mid \boldsymbol{l}]$ *is the conditional expectation of* a_s *when* $\boldsymbol{l}$ *is given,* $\mathscr{L} = \{\boldsymbol{l} \mid l_1 > \cdots > l_p > 0\}$, $\mathbf{I}$: $p \times p$ *dimensional unit matrix, c is a positive constant.*

Proof. We will prove

$$E_\Sigma[a_s \mid \boldsymbol{l}] \leqslant E_\Sigma[a_i \mid \boldsymbol{l}], \qquad \forall \boldsymbol{l} \in \mathscr{L},\ 1 \leqslant \forall s < \forall i \leqslant p.$$

The density of $\boldsymbol{l}$ and $\mathbf{H}$ with respect to Lebesgue measure and the invariant probability measure μ on $\mathcal{O}(p)$ (the group of p-dimensional orthogonal matrices) is given by

$$K \prod_{i=1}^{p} l_i^{(k-p-1)/2} \prod_{i<j} (l_i - l_j) \exp(-\tfrac{1}{2}\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{HDH}')),$$

where K is the normalizing constant. Therefore the conditional density of $\mathbf{H}$ given $\boldsymbol{l}$ is

$$f(\mathbf{H} \mid \boldsymbol{l}) = \frac{\exp(-(1/2)\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{HDH}'))}{\int_{\mathcal{O}(p)} \exp(-(1/2)\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{HDH}'))\, d\mu(\mathbf{H})}.$$

Now

$$\begin{aligned} \frac{E_\Sigma[a_s \mid \boldsymbol{l}]}{E_\Sigma[a_i \mid \boldsymbol{l}]} &= \frac{\int_{\mathcal{O}(p)} a_s \exp(-(1/2)\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{HDH}'))\, d\mu(\mathbf{H})}{\int_{\mathcal{O}(p)} a_i \exp(-(1/2)\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{HDH}'))\, d\mu(\mathbf{H})} \\ &= \frac{\int_{\mathcal{O}(p)} a_s \exp(-(1/2)\sum_{j=1}^{p} a_j l_j)\, d\mu(\mathbf{H})}{\int_{\mathcal{O}(p)} a_i \exp(-(1/2)\sum_{j=1}^{p} a_j l_j)\, d\mu(\mathbf{H})}. \end{aligned}$$

Hence $E_\Sigma[a_s \mid \boldsymbol{l}] \leqslant E_\Sigma[a_i \mid \boldsymbol{l}]$ if and only if

$$\int_{\mathcal{O}(p)} (a_s - a_i) \exp\left(-\frac{1}{2}\sum_{j=1}^{p} a_j l_j\right) d\mu(\mathbf{H}) \leqslant 0. \tag{2.2}$$

Also note that $E_\Sigma[a_s \mid \boldsymbol{l}] = E_\Sigma[a_i \mid \boldsymbol{l}]$ if and only if (2.2) holds with equality.

Now $\mu(\mathbf{H})$ is invariant with respect to permutation of columns of $\mathbf{H}$. Therefore by interchanging s and i, the left-hand side of (2.2) can be written as

$$\int_{\mathcal{O}(p)} (a_i - a_s) \exp\left(-\frac{1}{2} \sum_{j \neq s, i} a_j l_j - \frac{1}{2} a_i l_s - \frac{1}{2} a_s l_i \right) d\mu(\mathbf{H}). \tag{2.3}$$

Adding (2.3) and the left side of (2.2), $E_\Sigma[a_s \mid \boldsymbol{l}] \leqslant E_\Sigma[a_i \mid \boldsymbol{l}]$ if and only if

$$\int_{\mathcal{O}(p)} (a_s - a_i)[e^{-(a_i l_i + a_s l_s)/2} - e^{-(a_i l_s + a_s l_i)/2}] \times \exp\left(-\frac{1}{2} \sum_{j \neq s, i} a_j l_j \right) d\mu(\mathbf{H}) \leqslant 0. \tag{2.4}$$

Noting $l_s > l_i$ and $l_i a_i + l_s a_s - (l_i a_s + l_s a_i) = (l_s - l_i)(a_s - a_i)$, we have

$$\begin{aligned} e^{-(a_i l_i + a_s l_s)/2} &< e^{-(a_i l_s + a_s l_i)/2} \quad &\text{if} \quad a_s > a_i \\ e^{-(a_i l_i + a_s l_s)/2} &= e^{-(a_i l_s + a_s l_i)/2} \quad &\text{if} \quad a_s = a_i \\ e^{-(a_i l_i + a_s l_s)/2} &> e^{-(a_i l_s + a_s l_i)/2} \quad &\text{if} \quad a_s < a_i. \end{aligned}$$

Therefore the integrand of the left-hand side of (2.4) is always nonpositive and equal to zero if and only if $a_s = a_i$.

If $\mathbf{\Sigma} = c\mathbf{I}$, i.e., $\sigma_1^2 = \cdots = \sigma_p^2 = c$, then

$$a_i = \sum_{m=1}^{p} h_{mi}^2 \sigma_m^{-2} = c^{-1} \sum_{m=1}^{p} h_{mi}^2 = c^{-1} = a_s.$$

Hence

$$E_\Sigma[a_s \mid \boldsymbol{l}] = E_\Sigma[a_i \mid \boldsymbol{l}].$$

Now suppose $\mathbf{\Sigma} \neq c\mathbf{I}$, i.e.,

$$\sigma_{j_1}^2 > \sigma_{j_2}^2, \qquad 1 \leqslant \exists j_1 < \exists j_2 \leqslant p.$$

Choose such $\mathbf{H}^* = (h_{ij})$ from $\mathcal{O}(p)$ that satisfies

$$\begin{aligned} h_{j_1 i}^2 &= 1, \qquad h_{mi}^2 = 0 \quad (m \neq j_1) \\ h_{j_2 s}^2 &= 1, \qquad h_{ms}^2 = 0 \quad (m \neq j_2). \end{aligned}$$

For $\boldsymbol{H}^*$,

$$a_i = \sigma_{j_1}^{-2} < \sigma_{j_2}^{-2} = a_s.$$

There is a neighbourhood of $\mathbf{H}^*$ in $\mathcal{O}(p)$ (say $\mathcal{H}$) that satisfies

$$a_i < a_s \qquad \text{for} \quad \mathcal{H} \text{ and } \mu(\mathcal{H}) > 0.$$

Since the integrand (2.4) is negative on $\mathcal{H}$,

$$E_\Sigma[a_s \mid \mathbf{l}] < E_\Sigma[a_i \mid \mathbf{l}]. \quad \blacksquare$$

Let $\hat{\mathbf{\Sigma}} = \mathbf{H\Psi H}'$ as in (1.1), and modify $\hat{\mathbf{\Sigma}}$ as

$$\hat{\mathbf{\Sigma}}^0 = \mathbf{H\Psi}^0\mathbf{H}', \qquad \mathbf{\Psi}^0 = \operatorname{diag}(\psi_1^0(\mathbf{l}), ..., \psi_p^0(\mathbf{l})), \tag{2.5}$$

where $\psi_i^0(\mathbf{l})$ is the ith largest element in $(\psi_1(\mathbf{l}), ..., \psi_p(\mathbf{l}))$, i.e., $\psi_1^0 \geqslant \cdots \geqslant \psi_p^0$.

Proposition 1. *If $P_\Sigma[\mathbf{\Psi}^0 \neq \mathbf{\Psi}] > 0$ $\exists \mathbf{\Sigma}$, then $\hat{\mathbf{\Sigma}}^0$ dominates $\hat{\mathbf{\Sigma}}$.*

Proof. By (1.2),

$$\begin{aligned} L(\hat{\mathbf{\Sigma}}^0, \mathbf{\Sigma}) - L(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) &= \operatorname{tr}(\hat{\mathbf{\Sigma}}^0\mathbf{\Sigma}^{-1}) - \operatorname{tr}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) \\ &= \operatorname{tr}(\mathbf{\Psi}^0\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H}) - \operatorname{tr}(\mathbf{\Psi}\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H}) \\ &= \operatorname{tr}((\mathbf{\Psi}^0 - \mathbf{\Psi})\,\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H}) \\ &= \sum_{j=1}^p (\psi_j^0 - \psi_j)(\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H})_{jj} \\ &= \sum_{j=1}^p (\psi_j^0 - \psi_j)\, a_j. \end{aligned}$$

Therefore the difference of the risks can be written as

$$\begin{aligned} R(\hat{\mathbf{\Sigma}}^0, \mathbf{\Sigma}) - R(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma}) &= E_\Sigma\left[E_\Sigma\left[\sum_{j=1}^p (\psi_j^0 - \psi_j)\, a_j \mid \mathbf{l}\right]\right] \\ &= E_\Sigma\left[\sum_{j=1}^p E_\Sigma[(\psi_j^0 - \psi_j)\, a_j \mid \mathbf{l}]\right] \\ &= E_\Sigma\left[\sum_{j=1}^p (\psi_j^0 - \psi_j)\, E_\Sigma[a_j \mid \mathbf{l}]\right]. \end{aligned} \tag{2.6}$$

Hence it suffices to show

$$\sum_{j=1}^p (\psi_j^0 - \psi_j)\, E_\Sigma[a_j \mid \mathbf{l}] \leqslant 0, \qquad \forall \mathbf{l} \in \mathcal{L},\ \forall \mathbf{\Sigma},$$

where $\mathscr{L} = \{\boldsymbol{l} \mid l_1 > \cdots > l_p > 0\}$, and

$$\sum_{j=1}^{p} (\psi_j^0 - \psi_j)\, E_{\Sigma^*}[a_j \mid \boldsymbol{l}] < 0, \qquad \forall \boldsymbol{l} \in \{\boldsymbol{l} \mid \boldsymbol{\Psi}^0 \neq \boldsymbol{\Psi}\},\ \exists \boldsymbol{\Sigma}^*,$$

where $\boldsymbol{\Sigma}^*$ satisfies $P_{\Sigma^*}[\boldsymbol{\Psi}^0 \neq \boldsymbol{\Psi}] > 0$.

Using the Abel's identity,

$$\begin{aligned} c_1 d_1 + \cdots + c_p d_p = {} & (c_1 - c_2)\, d_1 + (c_2 - c_3)(d_1 + d_2) \\ & + \cdots + (c_{p-1} - c_p)(d_1 + \cdots + d_{p-1}) \\ & + c_p(d_1 + \cdots + d_p), \end{aligned}$$

we obtain

$$\begin{aligned} \sum_{j=1}^{p} (\psi_j^0 - \psi_j)\, E_\Sigma[a_j \mid \boldsymbol{l}] = {} & (\psi_1^0 - \psi_1)(E_\Sigma[a_1 \mid \boldsymbol{l}] - E_\Sigma[a_2 \mid \boldsymbol{l}]) \\ & + (\psi_1^0 + \psi_2^0 - \psi_1 - \psi_2)(E_\Sigma[a_2 \mid \boldsymbol{l}] - E_\Sigma[a_3 \mid \boldsymbol{l}]) \\ & + \cdots + (\psi_1^0 + \cdots + \psi_{p-1}^0 - \psi_1 - \cdots - \psi_{p-1}) \\ & \times (E_\Sigma[a_{p-1} \mid \boldsymbol{l}] - E_\Sigma[a_p \mid \boldsymbol{l}]). \end{aligned} \tag{2.7}$$

Note that $(\psi_1^0, ..., \psi_p^0)$ majorizes $(\psi_1, ..., \psi_p)$, i.e.,

$$\sum_{i=1}^{j} \psi_i^0 - \sum_{i=1}^{j} \psi_i \geqslant 0, \qquad 1 \leqslant \forall j \leqslant p-1 \text{ and } \sum_{i=1}^{p} \psi_i^0 = \sum_{i=1}^{p} \psi_i. \tag{2.8}$$

By Lemma 1,

$$E_\Sigma[a_i \mid \boldsymbol{l}] \leqslant E_\Sigma[a_{i+1} \mid \boldsymbol{l}], \qquad \forall \boldsymbol{l} \in \mathscr{L},\ \forall \boldsymbol{\Sigma},\ 1 \leqslant \forall i \leqslant p-1. \tag{2.9}$$

Equations (2.7), (2.8), (2.9) imply

$$\sum_{j=1}^{p} (\psi_j^0 - \psi_j)\, E_\Sigma[a_j \mid \boldsymbol{l}] \leqslant 0, \qquad \forall \boldsymbol{l} \in \mathscr{L},\ \forall \boldsymbol{\Sigma}.$$

Now by the continuity of $P_\Sigma[\boldsymbol{\Psi}^0 \neq \boldsymbol{\Psi}]$ with respect to $\boldsymbol{\Sigma}$,

$$P_{\Sigma^*}[\boldsymbol{\Psi}^0 \neq \boldsymbol{\Psi}] > 0, \qquad \exists \boldsymbol{\Sigma}^* \neq c\mathbf{I}. \tag{2.10}$$

By Lemma 1,

$$E_{\Sigma^*}[a_i \mid \boldsymbol{l}] < E_{\Sigma^*}[a_{i+1} \mid \boldsymbol{l}], \qquad \forall \boldsymbol{l} \in \mathscr{L},\ 1 \leqslant \forall i \leqslant p-1; \tag{2.11}$$

$\boldsymbol{\Psi}^0 \neq \boldsymbol{\Psi}$ implies

$$\sum_{i=1}^{j} \psi_i^0 > \sum_{i=1}^{j} \psi_i, \qquad 1 \leqslant \exists j \leqslant p-1. \tag{2.12}$$

Therefore, by (2.7), (2.8), (2.11), (2.12),

$$\sum_{j=1}^{p} (\psi_j^0 - \psi_j)\, E_{\Sigma^*}[a_j \mid \mathbf{l}] < 0, \qquad \forall \mathbf{l} \in \{\mathbf{l} \mid \Psi^0 \neq \Psi\}. \quad \blacksquare$$

We now consider another modification of non-order-preserving estimators, we modify

$$\hat{\mathbf{\Sigma}} = \mathbf{H}\Psi\mathbf{H}', \qquad \Psi = \operatorname{diag}(\psi_1, ..., \psi_p)$$

as

$$\hat{\mathbf{\Sigma}}^m = \mathbf{H}\Psi^m\mathbf{H}', \qquad \Psi^m = \operatorname{diag}(\psi_1^m, ..., \psi_p^m). \tag{2.13}$$

Before stating the modification, we define the function $M_{t+1}^t, ..., M_p^t$ $(t = 0, ..., p-1)$ as

$$M_l^t(\psi_1, ..., \psi_p) = \left(\sum_{j=t+1}^{l} \psi_j\right) \Big/ (l-t) \qquad (l = t+1, ..., p). \tag{2.14}$$

Now we state the algorithm of modification:

Step (1). Define $\mathcal{M}_1 = \{m \mid 1 \leqslant m \leqslant p,\ \ M_m^0 = \max_{1 \leqslant i \leqslant p} M_i^0\}$, $m_1 = \min_{m \in \mathcal{M}_1} m$.

Step (2). Define $\mathcal{M}_2 = \{m \mid m_1 + 1 \leqslant m \leqslant p,\ M_m^{m_1} = \max_{m_1+1 \leqslant i \leqslant p} M_i^{m_1}\}$, $m_2 = \min_{m \in \mathcal{M}_2} m$.

...

Step $(j+1)$. Define $\mathcal{M}_{j+1} = \{m \mid m_j + 1 \leqslant m \leqslant p,\ M_m^{m_j} = \max_{m_j+1 \leqslant i \leqslant p} M_i^{m_j}\}$, $m_{j+1} = \min_{m \in \mathcal{M}_{j+1}} m$.

...

This process ends when $m_s = p$. Then we obtain the sequence

$$m_0(=0) < m_1 < m_2 < \cdots < m_s = p.$$

Now we define ψ_i^m $(i = 1, ..., p)$ as

$$\psi_i^m = M_{m_j}^{m_{j-1}} = \left(\sum_{l=m_{j-1}+1}^{m_j} \psi_l\right) \Big/ (m_j - m_{j-1}),$$
$$\text{when} \quad m_{j-1} + 1 \leqslant i \leqslant m_j \ (j = 1, ..., s). \tag{2.15}$$

LEMMA 2. *The modified estimator* ψ_i^m $(i = 1, ..., p)$ *is order-preserving.*

Proof. Assume $\psi_i^m < \psi_{i+1}^m$, $1 \leqslant \exists i \leqslant p-1$. Then by (2.15),

$$\left(\sum_{l=m_{j-1}+1}^{m_j} \psi_l\right) \Big/ (m_j - m_{j-1}) < \left(\sum_{l=m_j+1}^{m_{j+1}} \psi_l\right) \Big/ (m_{j+1} - m_j), \qquad 1 \leqslant \exists j \leqslant s-1. \tag{2.16}$$

By the definition (2.14),

$$\begin{aligned} M_{m_j}^{m_{j-1}} &= \left(\sum_{l=m_{j-1}+1}^{m_j} \psi_l\right) \Big/ (m_j - m_{j-1}) \\ M_{m_{j+1}}^{m_{j-1}} &= \left(\sum_{l=m_{j-1}+1}^{m_{j+1}} \psi_l\right) \Big/ (m_{j+1} - m_{j-1}) \\ &= \left(\sum_{l=m_{j-1}+1}^{m_j} \psi_l + \sum_{l=m_j+1}^{m_{j+1}} \psi_l\right) \Big/ (m_{j+1} - m_{j-1}). \end{aligned}$$

By (2.16),

$$\left(\sum_{l=m_j+1}^{m_{j+1}} \psi_l\right) \Big/ (m_{j+1} - m_{j-1}) > \frac{m_{j+1} - m_j}{(m_j - m_{j-1})(m_{j+1} - m_{j-1})} \sum_{l=m_{j-1}+1}^{m_j} \psi_l.$$

Therefore,

$$\begin{aligned} M_{m_{j+1}}^{m_{j-1}} &> \left\{\frac{m_{j+1} - m_j}{(m_j - m_{j-1})(m_{j+1} - m_{j-1})} + \frac{1}{m_{j+1} - m_{j-1}}\right\} \sum_{l=m_{j-1}+1}^{m_j} \psi_l \\ &= \sum_{l=m_{j-1}+1}^{m_j} \psi_l / (m_j - m_{j-1}) \\ &= M_{m_j}^{m_{j-1}}. \end{aligned}$$

However, $M_{m_j}^{m_{j-1}} \geqslant M_{m_{j+1}}^{m_{j-1}}$ by the definitions of m_j and $M_{m_j}^{m_{j-1}}$. This is a contradiction. Hence,

$$\psi_i^m \geqslant \psi_{i+1}^m, \qquad 1 \leqslant \forall i \leqslant p-1. \quad \blacksquare$$

LEMMA 3. *$(\psi_1, ..., \psi_p)$ is majorized by $(\psi_1^m, ..., \psi_p^m)$, i.e.,*

$$\sum_{i=1}^{l} \psi_i \leqslant \sum_{i=1}^{l} \psi_i^m, \qquad 1 \leqslant \forall l \leqslant p-1, \qquad \sum_{i=1}^{p} \psi_i = \sum_{i=1}^{p} \psi_i^m.$$

Proof. Since

$$\sum_{i=m_{j-1}+1}^{m_j} \psi_i^m = \sum_{i=m_{j-1}+1}^{m_j} \psi_i, \qquad j=1, ..., s,$$

the equality

$$\sum_{i=1}^{p} \psi_i^m = \sum_{j=1}^{s} \sum_{i=m_{j-1}+1}^{m_j} \psi_i^m = \sum_{j=1}^{s} \sum_{i=m_{j-1}+1}^{m_j} \psi_i = \sum_{i=1}^{p} \psi_i$$

holds. Now suppose $m_{j-1}+1 \leqslant l \leqslant m_j$, $1 \leqslant \exists j \leqslant s$; then

$$\begin{aligned}\sum_{i=1}^{l} \psi_i^m &= \sum_{i=1}^{m_1} \psi_i^m + \sum_{i=m_1+1}^{m_2} \psi_i^m + \cdots + \sum_{i=m_{j-2}+1}^{m_{j-1}} \psi_i^m + \sum_{i=m_{j-1}+1}^{l} \psi_i^m \\ &= \sum_{i=1}^{m_{j-1}} \psi_i + \sum_{i=m_{j-1}+1}^{l} \psi_i^m.\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{i=1}^{l} \psi_i^m - \sum_{i=1}^{l} \psi_i &= \sum_{i=m_{j-1}+1}^{l} \psi_i^m - \sum_{i=m_{j-1}+1}^{l} \psi_i \\ &= (l-m_{j-1})\, M_{m_j}^{m_{j-1}} - \sum_{i=m_{j-1}+1}^{l} \psi_i \\ &\geqslant (l-m_{j-1})\, M_{l}^{m_{j-1}} - \sum_{i=m_{j-1}+1}^{l} \psi_i \\ &= \sum_{i=m_{j-1}+1}^{l} \psi_i - \sum_{i=m_{j-1}+1}^{l} \psi_i = 0. \quad \blacksquare\end{aligned}$$

Actually, ψ_i^m $(i=1, ..., p)$ is the isotonic regression of ψ_i $(i=1, ..., p)$; i.e.,

$$\min_{f \in \mathscr{F}} \sum_{i=1}^{p} (f_i - \psi_i)^2 = \sum_{i=1}^{p} (\psi_i^m - \psi_i)^2, \tag{2.17}$$

where $\mathscr{F} = \{f = (f_1, ..., f_p) \in \Re^p \mid f_1 \geqslant \cdots \geqslant f_p\}$. For the proof, see Theorem 1.2.1 of Robertson, Wright, and Dykstra [7]. For completeness we state another proof in the Appendix.

PROPOSITION 2. *If $P_\Sigma[\mathbf{\Psi}^m \neq \mathbf{\Psi}] > 0$ $\exists \Sigma$, then $\hat{\mathbf{\Sigma}}^m$ dominates $\hat{\mathbf{\Sigma}}$.*

Proof. As in Proposition 1,

$$L(\hat{\boldsymbol\Sigma}^m, \boldsymbol\Sigma) - L(\hat{\boldsymbol\Sigma}, \boldsymbol\Sigma) = \sum_{j=1}^{p} (\psi_j^m - \psi_j)\, a_j - \left(\sum_{j=1}^{p} \log \psi_j^m - \sum_{j=1}^{p} \log \psi_j \right),$$

$$\begin{aligned}
\sum_{j=1}^{p} \log \psi_j^m &= \sum_{j=1}^{s} (m_j - m_{j-1}) \log \left\{ \sum_{l=m_{j-1}+1}^{m_j} \psi_l/(m_j - m_{j-1}) \right\} \\
&= \sum_{j=1}^{s} \log \left\{ \sum_{l=m_{j-1}+1}^{m_j} \psi_l/(m_j - m_{j-1}) \right\}^{m_j - m_{j-1}} \\
&\geqslant \sum_{j=1}^{s} \log \prod_{l=m_{j-1}+1}^{m_j} \psi_l \\
&= \sum_{j=1}^{s} \sum_{l=m_{j-1}+1}^{m_j} \log \psi_l \\
&= \sum_{j=1}^{p} \log \psi_j.
\end{aligned}$$

Therefore,

$$L(\hat{\boldsymbol\Sigma}^m, \boldsymbol\Sigma) - L(\hat{\boldsymbol\Sigma}, \boldsymbol\Sigma) \leqslant \sum_{j=1}^{p} (\psi_j^m - \psi_j)\, a_j$$

and

$$\begin{aligned}
R(\hat{\boldsymbol\Sigma}^m, \boldsymbol\Sigma) - R(\hat{\boldsymbol\Sigma}, \boldsymbol\Sigma) &\leqslant E_\Sigma \left[\sum_{j=1}^{p} (\psi_j^m - \psi_j)\, a_j \right] \\
&= E_\Sigma \left[\sum_{j=1}^{p} (\psi_j^m - \psi_j)\, E_\Sigma [a_j \mid \boldsymbol l] \right].
\end{aligned}$$

By Lemma 3, $(\psi_1^m, ..., \psi_p^m)$ majorizes $(\psi_1, ..., \psi_p)$. Hence completely similarly as in the proof of Proposition 1, we have

$$\sum_{j=1}^{p} (\psi_j^m - \psi_j)\, E_\Sigma [a_j \mid \boldsymbol l] \leqslant 0, \qquad \forall \boldsymbol l \in \mathscr{L},\ \forall \boldsymbol\Sigma,$$

where $\mathscr{L} = \{\boldsymbol l \mid l_1 > \cdots > l_p > 0\}$, and

$$\sum_{j=1}^{p} (\psi_j^m - \psi_j)\, E_{\Sigma^*} [a_j \mid \boldsymbol l] < 0, \qquad \forall \boldsymbol l \in \{ l \mid \boldsymbol\Psi^m \neq \boldsymbol\Psi \},\ \exists \boldsymbol\Sigma^*,$$

where $\boldsymbol\Sigma^*$ satisfies $P_{\Sigma^*}[\boldsymbol\Psi^m \neq \boldsymbol\Psi] > 0$. ∎

We now consider the admissibility within the subfamily of estimators of the form (1.3). The unbiased estimator of $E_\Sigma[\operatorname{tr}(\hat{\Sigma}\Sigma^{-1}) - \log|\hat{\Sigma}|]$ given by Stein [2], Haff [3, 4] is

$$\hat{R}(\hat{\Sigma}, \Sigma) = \sum_{i=1}^{p} \left\{ (k-p-1)\frac{\psi_i(\boldsymbol{l})}{l_i} - \log \psi_i(\boldsymbol{l}) \right\} + 2\sum_{i>j} \frac{\psi_i(\boldsymbol{l}) - \psi_j(\boldsymbol{l})}{l_i - l_j} + 2\sum_{i=1}^{p} \frac{\partial \psi_i(\boldsymbol{l})}{\partial l_i}.$$

If we let $\psi_i = l_i d_i$, we have

$$\hat{R}(\hat{\Sigma}, \Sigma) = \sum_{i=1}^{p} \{\delta_i^{-1} d_i - \log(l_i d_i)\} + 2\sum_{i<j} \frac{l_j}{l_i - l_j}(d_i - d_j) + 2\sum_{i=1}^{p} l_i \frac{\partial d_i}{\partial l_i}.$$

Especially if d_i is constant,

$$\hat{R}(\hat{\Sigma}, \Sigma) = \sum_{i=1}^{p} \{\delta_i^{-1} d_i - \log(l_i d_i)\} + 2\sum_{i<j} \frac{l_j}{l_i - l_j}(d_i - d_j).$$

COROLLARY 1. *The estimators of the form* (1.3) *are inadmissible.*

Proof. Fix d_i $(i = 1, ..., p)$ in (1.3) and let $\hat{\Sigma}_d$ denote the estimator with these d_i's. Let $\hat{\Sigma}_\delta$ be the estimator given in (1.4). The risk difference can be written as

$$R(\hat{\Sigma}_d, \Sigma) - R(\hat{\Sigma}_\delta, \Sigma) = E_\Sigma[\hat{R}(\hat{\Sigma}_d, \Sigma) - \hat{R}(\hat{\Sigma}_\delta, \Sigma)].$$

Now

$$\hat{R}(\hat{\Sigma}_d, \Sigma) - \hat{R}(\hat{\Sigma}_\delta, \Sigma) = \sum_{i=1}^{p} [\{\delta_i^{-1} d_i - \log(l_i d_i)\} - \{\delta_i^{-1}\delta_i - \log(l_i \delta_i)\}] + 2\sum_{i<j} \frac{l_j}{l_i - l_j}\{(d_i - d_j) - (\delta_i - \delta_j)\}.$$

For fixed l_i, $i = 1, ..., p$, $f(x) = \delta_i^{-1} x - \log(l_i x)$ is minimized by $x = \delta_i$. Therefore,

$$\delta_i^{-1} d_i - \log(l_i d_i) \geqslant \delta_i^{-1}\delta_i - \log(l_i \delta_i).$$

Now suppose $d_1 \geqslant \cdots \geqslant d_p$, then

$$\sum_{i<j} \frac{l_j}{l_i - l_j} \{(d_i - d_j) - (\delta_i - \delta_j)\} > 0,$$

since $\delta_1 < \cdots < \delta_p$. This implies that if $d_1 \geqslant \cdots \geqslant d_p$, then $\hat{R}(\hat{\Sigma}_d, \Sigma) > \hat{R}(\hat{\Sigma}_\delta, \Sigma)$ and, hence, $\hat{\Sigma}_d$ is dominated by $\hat{\Sigma}_\delta$. Therefore, if $\hat{\Sigma}_d$ is not dominated by $\hat{\Sigma}_\delta$, it must be

$$d_i < d_j, \qquad \exists i < \exists j.$$

Therefore if $d_i/d_j < l_j/l_i$,

$$\psi_i = l_i d_i < l_j d_j = \psi_j.$$

This means $\hat{\Sigma}_d$ must be non-order-preserving if it is not dominated by $\hat{\Sigma}_\delta$. But Proposition 1 states that non-order-preserving estimator is inadmissible. ∎

As we have shown above, order-preserving estimators form a complete class. Hence we should look for a good estimator in the class of order-preserving estimators. We considered modification of good non-order-preserving estimators. However, we may restrict our attention to the class of order-preserving estimators from the beginning and try to find good estimators in this class. This approach is taken in Haff [8] and Perron [9].

3. Simulation Study

A Monte Carlo simulation study was performed to compute the risks of six estimators, i.e., (1.4), (1.5), and modifications of these estimators using order statistics and using isotonic regression. The dimension p and the degree of freedom k ranged from 2 to 10, 10 to 20, and we considered several typical Σ's. Summary of the simulation study is as follows:

1. The improvement of modified estimators over original estimators (1.4), (1.5) was at most 0.7%.

2. The improvement of modified estimators over original estimators decreases as k increases.

3. It depends on Σ which of the two modifications is better. Roughly speaking, modification using isotonic regression is better when Σ is close to I. On the other hand, modification using order statistics is better when the eigenvalues of Σ are far away from each other. This can be naturally conjectured from the proof of Propositions 1 and 2 and (2.1).

TABLE I

σ_1^2	σ_2^2	σ_3^2	R1	R2	S	S-O	S-I	D	D-O	D-I
1.00	1.00	1.00	7444	7444	2.414	2.414	2.404	2.414	2.414	2.404
						0.00%	0.40%	0.00%	0.00%	0.40%
1.00	0.10	0.10	8392	8392	2.627	2.624	2.618	2.627	2.624	2.618
						0.10%	0.33%	0.00%	0.10%	0.33%
1.00	0.01	0.01	8634	8634	2.728	2.728	2.721	2.728	2.728	2.721
						0.00%	0.27%	0.00%	0.00%	0.27%
1.00	0.50	0.10	8432	8432	2.646	2.633	2.634	2.646	2.633	2.634
						0.49%	0.45%	0.00%	0.49%	0.45%
1.00	0.10	0.01	9361	9361	2.879	2.868	2.871	2.879	2.868	2.871
						0.41%	0.29%	0.00%	0.41%	0.29%

Here we present a result for the case $p=3$ and $k=3$ in Table I. We used 10,000 randomly generated Wishart matrices. In Table I,

1. $\Sigma = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$
2. R1. Number of times (out of 10,000), where ψ_i $(i=1, ..., p)$ of (1.4) was in descending order.
3. R2. Number of times (out of 10,000), where ψ_i $(i=1, ..., p)$ of (1.5) was in descending order.
4. S. The risk of (1.4).
5. S-O(I). The risk of modified estimator of (1.4) using order statistics (isotonic regression).
6. D. The risk of (1.5).
7. D-O(I). The risk of modified estimator of (1.5) using order statistics (isotonic regression).
8. The figures under risks show the improvement over (1.4).

Appendix

LEMMA 4. *Equation* (2.15) *is the* (*unique*) *solution of* (2.17).

Proof. It suffices to show

$$\sum_{i=1}^{p} (\psi_i - \psi_i^m)(\psi_i^m - f_i) \geqslant 0 \qquad \forall f = (f_1, ..., f_p) \in \mathscr{F} \tag{A.1}$$

(see Theorem 1.3.1 of Robertson, Wright, and Dykstra [7]);

$$\begin{aligned}\sum_{i=1}^{p} (\psi_i-\psi_i^m)\,\psi_i^m &= \sum_{i=1}^{p} \psi_i\psi_i^m - \sum_{i=1}^{p} \psi_i^m\psi_i^m \\ &= \sum_{j=1}^{s} \sum_{l=m_{j-1}+1}^{m_j} \psi_l\psi_l^m - \sum_{j=1}^{s} \sum_{l=m_{j-1}+1}^{m_j} \psi_l^m\psi_l^m \\ &= \sum_{j=1}^{s} M_{m_j}^{m_{j-1}} \sum_{l=m_{j-1}+1}^{m_j} \psi_l - \sum_{j=1}^{s} M_{m_j}^{m_{j-1}} \sum_{l=m_{j-1}+1}^{m_j} \psi_l^m \\ &= \sum_{j=1}^{s} M_{m_j}^{m_{j-1}} \left(\sum_{l=m_{j-1}+1}^{m_j} \psi_l - \sum_{l=m_{j-1}+1}^{m_j} \psi_l^m \right) \\ &= 0.\end{aligned}$$

Hence, it suffices to show

$$\sum_{i=1}^{p} (\psi_i-\psi_i^m) f_i \leqslant 0 \qquad \forall \in \mathscr{F}. \tag{A.2}$$

Using Abel's identity,

$$\begin{aligned}\sum_{i=1}^{p} (\psi_i-\psi_i^m) f_i &= (\psi_1-\psi_1^m)(f_1-f_2)+(\psi_1+\psi_2-\psi_1^m-\psi_2^m)(f_2-f_3) \\ &\quad + \cdots + (\psi_1+\cdots+\psi_{p-1}-\psi_1^m-\cdots-\psi_{p-1}^m)(f_{p-1}-f_p).\end{aligned}$$

Note $\sum_{i=1}^{p} \psi_i = \sum_{i=1}^{p} \psi_i^m$ by Lemma 3. Since $f_j \geqslant f_{j+1}$ $(j=1, ..., p-1)$ and Lemma 3, (A.2) is obvious. ∎

Acknowledgments

We thank Professor K. Takeuchi and Dr. Kubokawa for their valuable suggestions.

References

[1] James, W., and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings, Berkeley Symp. Math. Statist. Probab.*, Vol. 1, pp. 361–380. Univ. of California Press, Berkeley.

[2] Stein, C. (1977). Unpublished notes on estimating the covariance matrix.

[3] Haff, L. R. (1977). Minimax estimators for a multinormal precision matrix. *J. Multivariate Anal.* **7** 374–385.

[4] Haff, L. R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.* **9** 531–542.

[5] Dey, D. K., and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. *Ann. Statist.* **13** 1581–1591.

[6] Lin, S. P., and Perlman, M. D. (1985). A Monte Carlo comparison of four estimators for a covariance matrix. In *Multivariate Analysis* (P. R. Krishnaiah, Ed.), Vol. 6, pp. 411–429. North Holland, Amsterdam.

[7] Robertson, T., Wright, F. T., Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, Wiley, New York.

[8] Haff, L. R. (1988). The variational form of certain Bayes estimators, personal communication.

[9] Perron, F. (1991). Minimax estimators of a covariance matrix, personal communication.